On the Capacity of an Ensemble of Channels with Differing Parameters

By E. A. WALVICK

(Manuscript received November 17, 1969)

To provide a mathematical tool for the evaluation of cable pairs, this paper suggests a quality measure which is based on information theory. While a group of cable pairs (paired wires) with a given gauge and construction are nominally equivalent, manufacturing tolerances and differences in installation and environment lead to variation from the nominal parameters.

To measure the quality of a group of channels (for example, cable pairs leaving a central office), this paper recommends the following procedure. Choose a fraction p of the original group and evaluate the mutual information between input and output of each channel in the subgroup, subject to the input to each channel being chosen from the same process. Then, by choosing the proper process, maximize the smallest mutual information in the subgroup. This largest possible minimum mutual information is a quality measure for the subgroup. Next, apply this measure to all subgroups of fractional size p; the subgroup with the highest measure provides the numerical value of the quality measure for the original group relative to fraction p. Repeat this procedure for all p $(0 \leqq p \leqq 1)$. The resulting function is the suggested quality measure of the group.

To illustrate the above measure of quality, we derive the capacity of an ensemble of channels with stationary Gaussian inputs, additive noise, and crosstalk. In the Appendix we derive the capacity of a single such channel.

I. INTRODUCTION

Cables are usually analyzed as if all of the components had a particular set of parameters (for example, nominal, worst case, and so on). Because of manufacturing tolerances, installation differences and various environmental effects, however, transmission parameters actually vary from pair to pair. To account for these variations, this paper takes an

approach based on information theory. We consider the cable network to be a statistical population of channels which have parameters that vary from channel to channel. We propose a quality measure for the network based on this model.

After defining channel capacity, we present the suggested capacity definition for a group of channels (as outlined in the Abstract). Using this definition, we provide an example in Section IV in which the capacity of a group of channels with stationary Gaussian inputs, additive noise and crosstalk is derived. The capacity for a single such channel is found in the Appendix. In Section 4.1 the crosstalk is assumed to differ from channel to channel; in Section 4.2 the channel attenuation is assumed variable; and in Section 4.3 both the crosstalk and attenuation vary from channel to channel.

These results indicate the trade-off between design rate for a transmission system and the expected fraction of channels which will be capable of error-free transmission at the design rate.

This technique can also be used to evaluate different parameter distributions as may result from tighter production controls.

II. CHANNEL CAPACITY

A channel is defined as a probabilistic mapping of one stochastic process onto another (for our problem we consider the processes to be time functions). (See Fig. 1.)

Let

$s(t)$ be the input stochastic process,
$r(t)$ be the output stochastic process,
$s_T = \{s(t) : t \in [-T/2, T/2]\}$,
s be s_∞ .

Then the operation of the channel can be written in terms of a probabilistic mapping F as

$$F\{s\} = r. \tag{1}$$

The capacity is defined as the maximum (over input processes) mutual information between input and output, that is,

Fig. 1 — Channel model.

$$C \equiv \sup_{s} \limsup_{T\to\infty} \frac{1}{T} I(r_T , s_T) = \sup_{s} \langle I(r, s)\rangle \tag{2}$$

where

$$\langle I(r, s)\rangle \equiv \limsup_{T\to\infty} \frac{1}{T} I(r_T , s_T). \tag{3}$$

$I(r_T , s_T)$ is the mutual information* between s_T and r_T
the supremum is taken over all possible distributions of input signals subject to some constraint (for example, fixed power),
and for a large class of channels including memoryless cbannels and colored Gaussian channels C is the maximum information rate (that is, the maximum error free transmittable rate).

The maximization of equation (2) will yield not only C, but more importantly perhaps, the properties of $s(t)$ which will achieve C.

III. CAPACITY DEFINITIONS FOR A GROUP OF CHANNELS

Now, consider the extension of the capacity definition to a group of channels. Capacity is dependent not only on the nature of the channel, but also the nature of the constraints placed on the input. For different sets of input constraints, different capacities will be obtained. This section contains two possible alternate capacity definitions [equations (5) and (6)] followed by the recommended definition [equation (10)].

A natural extension of equation (2) to a class of channels (formally tbe set $\{\omega : \omega \in \Omega\}$)

$$F^{(\omega)}\{s\} = r^{(\omega)}, \qquad \omega \in \Omega$$

(that is, $F^{(\omega)}$ is the mapping corresponding to cbannel ω) would be to define the capacity as the sum of the individual capacities, or the average capacity for an infinite set Ω. Tbat is, the capacity of each cbannel is:

$$C^{(\omega)} = \sup_{s^{(\omega)}} \langle I(r^{(\omega)}, s^{(\omega)})\rangle,$$

wbere the supremum is performed for each cbannel separately, constrained as before. Then one measure of the capacity of tbe ensemble could be the total capacity (for a finite set):

$$C_T = \sum_{\omega \in \Omega} C^{(\omega)}. \tag{4}$$

Another measure could be the average per-channel capacity

* See for example Gallager.[1]

$$\langle C \rangle_T = E\{C^{(\omega)}\}, \tag{5}$$

where E denotes the expectation.

As for the single channel, equation (4) [or (5)] yields both C_T (or $\langle C \rangle_T$) and the properties of the set $S = \{s^{(\omega)} : \omega \in \Omega\}$ which will achieve C_T (or $\langle C \rangle_T$). This number defines the maximum transmittable rate when the input processes are chosen for each channel individually. In many instances this may not be a practical measure in that it may be desirable to use a single signaling set on all members of Ω. A measure using a single signaling set has been suggested in the literature:[2]

$$C_B = \sup_s \inf_{\omega \in \Omega} \langle I(r^{(\omega)}, s) \rangle. \tag{6}$$

The desirable property of C_B is that it will result in a signal distribution which, when applied to any member, $\omega \in \Omega$, will permit transmission at rates arbitrarily close to C_B with arbitrarily small probability of error. That is, C_B is the maximum rate which will work on all members of the group when one process is sent over all channels. However, this seems to be an overly pessimistic measure in that if Ω should have even one member with poor transmission properties, C_B will reflect this single poor member in exactly the same way as if all of Ω were equally bad.

To overcome this difficulty a new capacity definition is introduced. This definition is actually a function rather than a single number for the group of channels. This definition is essentially C_B , restricted to the best subset, of size p, of the original group of channels, as a function of p. To formalize this notion:

Let $\Omega_\lambda(p)$ be a subset of Ω (indexed by λ) of fractional size p. That is, for $\omega \in \Omega$

$$\Pr\,\{\omega : \omega \in \Omega_\lambda(p)\} = p. \tag{7}$$

Let $\Omega(p)$ be the set of all such subsets:

$$\Omega(p) = \{\Omega_\lambda(p)\}. \tag{8}$$

Find C_B for each subset $\Omega_\lambda(p)$:

$$C_B[\Omega_\lambda(p)] = \sup_s \inf_{\omega \in \Omega_\lambda(p)} \langle I(r^{(\omega)}, s) \rangle. \tag{9}$$

Finally consider the supremum over all such subsets, that is

$$\begin{aligned} C(p) &\equiv \sup_\lambda \sup_s \inf_{\omega \in \Omega_\lambda(p)} \langle I(r^{(\omega)}, s) \rangle \\ &= \sup_\lambda C_{B\Omega_\lambda}. \end{aligned} \tag{10}$$

Note that equation (10) can also be written as:

$$C(p) = \sup_{s} \sup_{\lambda} \inf_{\omega \in \Omega_\lambda(p)} \langle I(r^{(\omega)}, s) \rangle.$$

The following coding theorem can be proved almost by inspection: $C(p)$ is the supremum of rates which can be transmitted error free over at least the fraction p of the original ensemble of channels when the inputs to all channels are from the same signal distribution. (Clearly, for larger p, more channels are required to be capable of error free transmission at rates arbitrarily close to $C(p)$ than for smaller p. Therefore, $C(p)$ decreases as p increases, more of the set of poor channels included.)

$C(p)$ would then be plotted as in Fig. 2 which intentionally represents an ensemble for which most of the channels have near nominal parameters, a small percentage have worse parameters, and a small percentage better parameters. From the figure, if all channels must have error-free transmission, then the design rate for any system can be no greater than C_B. However, if design criterion only requires p_1 of the channels to be error free then rate C_1 can be used. Similarly, if only p_2 of the channels need operate without errors, rate C_2 can be used.

$C(p)$ is a useful measure for the following reasons:

(*i*) If it is desired that a given fraction of the channels have satisfactory transmission, then $C(p)$ indicates the maximum permissible rate.

(*ii*) If the objective is to provide a given transmission rate $C(p)$, then the value of p indicates what percentage of the channels will be capable of operating without errors.

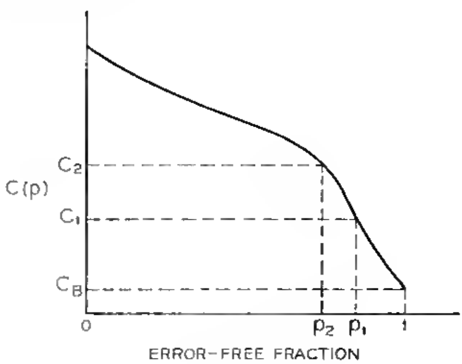

Fig. 2 — $C(p)$ vs. p.

(iii) When deciding between two alternative groups of channels, $C(p)$ can indicate those transmission rates for which one type is better than the other.

IV. EXAMPLE

To provide an example for tbe use of the $C(p)$ function, consider an ensemble of channels, each of which can be modeled as shown in Fig. 3.*

where

$H(\omega)$ is the channel transfer function,
$X(\omega)$ is the crosstalk transfer function,
$s(t)$ is the signal with one-sided power spectral density $S_S(\omega)$,
$\hat{s}(t)$ is the signal on an adjacent channel with the same power spectral density, and
$n(t)$ is the noise with power spectral density $S_N(\omega)$.

If only stationary Gaussian inputs are considered the average mutual information between the input and output of any channel in the ensemble is:[3,4]

$$\langle I(r, s)\rangle = -\frac{1}{2\pi}\int_0^\infty \log\left(1 - \frac{|S_{SR}(\omega)|^2}{S_R(\omega)S_S(\omega)}\right) d\omega, \tag{11}$$

where

$S_S(\omega)$ is the input signal power spectral density,

$$S_R(\omega) = S_S(\omega)\,|H(\omega)|^2\,(1 + |X(\omega)|^2) + S_N(\omega), \tag{12}$$

is the output power spectral density, and

$$S_{SR}(\omega) = S_S(\omega)H^*(\omega), \tag{13}$$

is the input-output cross power spectral density.
Then,

$$\langle I(r, s)\rangle = -\frac{1}{2\pi} \cdot\int_0^\infty \log\left[1 - \frac{S_S^2(\omega)\,|H(\omega)|^2}{S_S^2(\omega)\,|H(\omega)|^2\,(1 + |X(\omega)|^2) + S_S(\omega)S_N(\omega)}\right] d\omega. \tag{14}$$

---

* ω is used here to represent frequency and not probability spaces as in the first part of this paper.

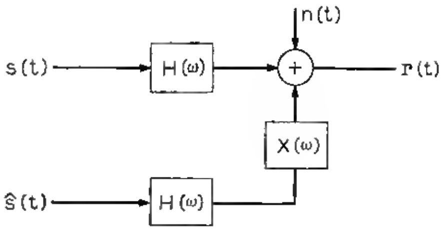

Fig. 3 — Channel model.

Define

$$\mathcal{S}_N(\omega) = \frac{S_N(\omega)}{\mid H(\omega) \mid^2}. \tag{15}$$

Then

$$\langle I(r, s) \rangle = \frac{1}{2\pi} \int_0^\infty \log \left[1 + \frac{S_S(\omega)}{\mid X(\omega) \mid^2 S_S(\omega) + \mathcal{S}_N(\omega)} \right] d\omega. \tag{16}$$

4.1 *Fixed Noise and Channel, Distributed Crosstalk*

Assume that the crosstalk parameters of all of the cables are the same except for a multiplier, that is:

$$\mid X_\xi(\omega) \mid = \epsilon_\xi \mid X(\omega) \mid. \tag{17}$$

Equation (16) indicates that for any choice of $S_S(\omega)$, $\langle I(r, s) \rangle$ decreases as ϵ_ξ increases. Thus, while equation (9) requires a minimization followed by a maximization, it is clear in the case that the "inf" for a given p occurs for the largest ϵ_ξ in the subset $\Omega_\lambda(p)$. Further, the "sup" is achieved using the spectrum that achieves capacity on the channel with the largest ϵ_ξ . Finally, the "sup" in equation (10) is achieved by choosing the $\Omega_\lambda(p)$ such that $0 \leq \epsilon \leq \epsilon_p$ where ϵ_p is such that

$$p = \int_0^{\epsilon_p} p_{\epsilon_\xi}(\epsilon) \, d\epsilon, \tag{18}$$

where $p_{\epsilon_\xi}(\epsilon)$ is the probability density of ϵ_ξ . Thus using the capacity result obtained in the Appendix

$$S_0(\omega) = [\mathcal{S}_N(\omega_{\max}) - \mathcal{S}_N(\omega)][1 - \tfrac{1}{4}\epsilon_p^2 \mid X(\omega) \mid^2 (\epsilon_p^2 \mid X(\omega) \mid^2 + 1) + \cdots]. \tag{19}$$

The capacity function can now be found:

$$C(p) = \frac{1}{2\pi}\left\{\int_0^{\omega_{max}} \left[\log\left[1 + \frac{1}{\epsilon_p^2 \mid X(\omega) \mid^2}\right]\right.\right.$$
$$\left.\left. + \tfrac{1}{2}\log\left\{1 + 4\epsilon_p^2 \mid X(\omega) \mid [\epsilon_p^2 \mid X(\omega) \mid^2 + 1] \frac{S_N(\omega_{max})}{S_N(\omega)}\right\}\right] d\omega\right\}. \tag{20}$$

Consider the following example to illustrate the above:

Let $S_N(\omega)$ be zero. Then from equation (16) the mutual information is independent of the signal spectrum and

$$C(p) = \frac{1}{2\pi}\int_0^{\infty} \log\left[1 + \frac{1}{\epsilon_p^2 \mid X(\omega) \mid^2}\right] d\omega. \tag{21}$$

(This is achieved for any spectrum with finite power which is strictly greater than zero for all frequencies.) Let

$$\mid X(\omega) \mid^2 = \omega^2 \tag{22}$$

and let $\log \epsilon_p$ be normally distributed with mean -8.2 and $\sigma = 0.17$. (The crosstalk figures are idealizations of typical figures for 22 gauge PIC). Then

$$p = \operatorname{erf}\left\{\frac{\log \epsilon_p + 8.2}{0.17}\right\} + 0.5 \tag{23}$$

where

$$\operatorname{erf}(x) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^x e^{-y^2/2}\, dy, \tag{24}$$

and

$$C(p) = \frac{1}{2\epsilon_p \ln 2} \quad \text{(in bits)}$$
$$= \frac{1}{2\epsilon_p} \quad \text{(in nats)}. \tag{25}$$

These results are indicated in Fig. 4. For $p = 0.8$, $C(p)$ is 6.3×10^7. For $p = 0.1$, $C(p)$ is 1.3×10^8. Thus a factor of two is realized in capacity by reducing the required usable fraction of channels by a factor of eight.

4.2 *Fixed Crosstalk and Noise, Distributed Channel*

Assume

$$S_N(\omega) = \epsilon_N\ , \tag{26}$$

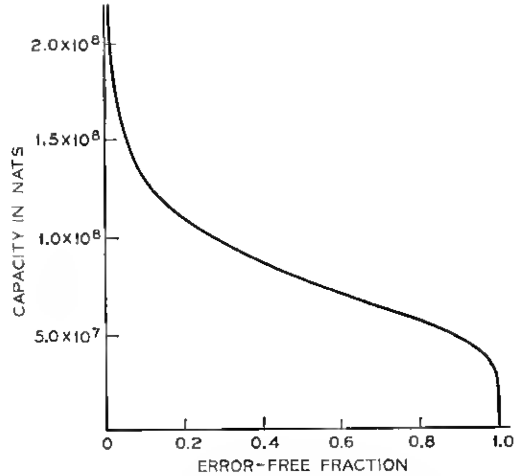

Fig. 4 — Capacity vs. fraction of satisfactory channels.

$$| H_\xi(\omega) | = e^{-\alpha_\xi(\omega)}, \tag{27}$$

and that $\alpha_\xi(\omega)$ is such that

$$\alpha_{\xi_1}(\omega_i) > \alpha_{\xi_2}(\omega_i) \Rightarrow \alpha_{\xi_1}(\omega_j) \geqq \alpha_{\xi_2}(\omega_j). \tag{28}$$

That is, if the attenuation of channel ξ_1 is greater than the attenuation of ξ_2 at one frequency, it is greater at all frequencies. Then, again design for the worst case, that is,

$$p = \int_0^{\alpha_p} p_{\alpha_\xi}(\alpha)\, d\alpha. \tag{29}$$

(Again the smaller α, the "better" the channel.)

$$S_0(\omega) = \epsilon_N[e^{2\alpha_p(\omega_{max})} - e^{2\alpha_p(\omega)}][1 - \tfrac{1}{4} | X(\omega) |^2 (| X(\omega) |^2 + 1)], \tag{30}$$

and

$$C(p) = \frac{1}{2\pi}\left\{\int_0^{\omega_{max}} \left\{\log\left[1 + \frac{1}{| X(\omega) |^2}\right]\right.\right.$$
$$\left.\left. + \tfrac{1}{2}\log\left[1 + 4 | X(\omega) |^2 (| X(\omega) |^2 + 1)\frac{e^{2\alpha_p(\omega_{max})}}{e^{2\alpha_p(\omega)}}\right]\right\} d\omega\right\}. \tag{31}$$

4.3 *Fixed Noise, Distributed Channel and Crosstalk*

Proceeding as in the previous two sections, choose the signal spectrum to achieve capacity on a channel with a particular ϵ_{p_s} and α_{p_s},

that is:

$$S_{0_x}(\omega) = \epsilon_N[\exp[2\alpha_{p_x}(\omega_{\max})] - \exp[2\alpha_{p_x}(\omega)]] \cdot [1 - \tfrac{1}{4}\epsilon_{p_x}^2 \mid X(\omega) \mid^2 (\epsilon_{p_x}^2 \mid X(\omega) \mid^2 + 1)], \tag{32}$$

$$C_x(p) = \frac{1}{2\pi}\left\{\int_0^{\omega_{\max}} \left\{\log\left(1 + \frac{1}{\epsilon_{p_x}^2 \mid X(\omega) \mid^2}\right) + \tfrac{1}{2}\log\left[1 + 4\epsilon_{p_x}^2 \mid X(\omega) \mid^2 (\epsilon_{p_x}^2 \mid X(\omega) \mid^2 + 1) \cdot \frac{\exp[2\alpha_{p_x}(\omega_{\max})]}{\exp[2\alpha_{p_x}(\omega)]}\right]\right\} d\omega\right\}. \tag{33}$$

Here, p is found by integrating the joint density of α_ξ and ϵ_ξ over the region where the mutual information is greater than $C_x(p)$. That is,

$$p = \iint_R p_{\epsilon_\xi, \alpha_\xi}(\epsilon, \alpha)\, d\epsilon\, d\alpha \tag{34}$$

where R is the region in the (ϵ, α) space, where

$$\frac{1}{2\pi}\int_0^{\omega_{\max}} \log\left[1 + \frac{1}{\epsilon_\xi \mid X(\omega) \mid^2 + \epsilon_N \exp 2\alpha_\xi(\omega)}\right] d\omega \geqq C_x(p). \tag{35}$$

The difficulty with this problem at this point is that there is no unique ϵ_{p_x} and α_{p_x} which yield a given p in equation (34). Thus it is necessary to search all such sets $(\epsilon_{p_x}, \alpha_{p_x})$ which yield the same p, and choose the one which yields the largest $C_x(p)$. While this may seem difficult, it can be implemented with a computer search.

V. CONCLUDING REMARKS

The quality measure suggested herein should be considered part of the development of a technique for comparison of groups of cables.

It should be pointed out that, while the definition of capacity for a single channel is straightforward, the definition of capacity for an ensemble of channels depends on certain assumptions concerning the use of the channels.

$\langle C \rangle_T$, the average per channel capacity, [equation (5)] assumes each channel will be used individually, and optimized individually. C_B, Blackwell's definition, [equation (6)] assumes the same input distributions will be used for all channels in the group and that the channels are used individually. Further, no more than the minimum rate transmittable over the worst channel will be sent over each channel.

The recommended definition, $C(p)$, equation (12), is similar to C_B except that the fraction $1 - p$ of the worst channels will not work satisfactorily and could be discarded.

The example carried out in Section IV uses a new result on the capacity of a single channel with crosstalk and additive Gaussian noise which is derived in the Appendix. The discussion on the calculus of variations, contained before equation (40), can be used to rigorously prove some old results for optimum spectra in the presence of additive Gaussian noise [see footnote, following equation (52)].

In applying $C(p)$, it is noted that when only one parameter is unknown (for example, the magnitude of the crosstalk), the capacity of the group of channels as a function of p (the fraction of usable channels) is a simple calculation. However, when more than one parameter is unknown (for example, the magnitude of the crosstalk as well as the channel attenuation), search techniques are indicated.

APPENDIX

Capacity Calculation for a Single Channel

Consider the channel model of Fig. 3. The mutual information expression is given in equation (16) and is repeated below:

$$\langle I(r, s)\rangle = \frac{1}{2\pi}\int_0^\infty \log\left[1 + \frac{S_S(\omega)}{\mid X(\omega)\mid^2 S_S(\omega) + \mathcal{S}_N(\omega)}\right]d\omega. \tag{36}$$

Now capacity for this channel is defined as

$$C = \sup_{S_S(\omega)} \langle I(r, s)\rangle, \tag{37}$$

subject to

$$-\frac{1}{2\pi}\int_0^\infty S_S(\omega)\, d\omega \leqq P = \text{power},$$

and

$$S_S(\omega) \geqq 0 \qquad \text{for all } \omega. \tag{38}$$

The method of variational calculus can be applied to determine the optimum solution.

Let

$S_0(\omega)$ be the assumed optimum solution, and
$\delta\cdot\epsilon(\omega)$ be a small perturbation, then
$S_S(\omega) = S_0(\omega) + \delta\cdot\epsilon(\omega)$.

In order to account for $S_S(\omega) \geqq 0$, $\delta \cdot \epsilon(\omega)$ must be nonnegative whenever $S_0(\omega) = 0$. That is, in order that $S_S(\omega)$ be a power spectral density, whenever $S_0(\omega) = 0$, the perturbation at that frequency must be such that the resulting density at that frequency remain nonnegative. Note that since $S_0(\omega)$ is to be the optimum solution, $\langle I(r, s) \rangle$ must be a maximum at $\delta = 0$ *for all permissible* $\delta \cdot \epsilon(\omega)$. This implies that $\partial \langle I \rangle / \partial \delta \mid_{\delta=0}$ will be negative for all permissible $\delta \cdot \epsilon(\omega)$ which approach zero from the positive side $(\delta \cdot \epsilon(\omega) \to 0^+)$, and $\partial \langle I \rangle / \partial \delta \mid_{\delta=0}$ will be positive for $\delta \cdot \epsilon(\omega) \to 0^-$. Or finally, $\partial \langle I \rangle / \partial \delta \mid_{\delta=0} = 0$ whenever $\delta \cdot \epsilon(\omega)$ can approach zero from either the positive or negative side and the derivative is defined. Now,

$$2\pi \frac{\partial \langle I \rangle}{\partial \delta} = \frac{\partial}{\partial \delta} \left[\int_0^{+\infty} \log \left[1 + \frac{1}{\mid X(\omega) \mid^2 + \dfrac{\mathcal{S}_N(\omega)}{S_0(\omega) + \delta \cdot \epsilon(\omega)}} \right] d\omega \right] + \frac{\partial}{\partial \delta} \left[-\mu \left\{ \int_0^{+\infty} [S_0(\omega) + \delta \cdot \epsilon(\omega)] \, d\omega - 2\pi P \right\} \right], \tag{39}$$

where μ is a Lagrangian multiplier used to introduce the power constraint.

Or

$$\left. \frac{\partial \langle I \rangle}{\partial \delta} \right|_{\delta=0} = \int_0^{\infty} \epsilon(\omega) \left[\frac{\mathcal{S}_N(\omega)}{A(\omega) S_0^2(\omega) + B(\omega) S_0(\omega) + \mathcal{S}_N^2(\omega)} - \mu \right] d\omega, \tag{40}$$

where

$$A(\omega) = \mid X(\omega) \mid^2 (\mid X(\omega) \mid^2 + 1),$$

$$B(\omega) = [2 \mid X(\omega) \mid^2 + 1] \, \mathcal{S}_N(\omega).$$

In equation (20) one now considers all permissible $\epsilon(\omega)$. Whenever $S_0(\omega)$ is nonzero (that is, $\delta \cdot \epsilon(\omega)$ is unrestricted), the integrand is zero, since for these frequencies, $\epsilon(\omega)$ is arbitrary. However, when $S_0(\omega)$ is zero $(\delta \cdot \epsilon(\omega) \geqq 0)$ all that can be said is that the integrand is *negative*. That is, as one approaches the boundary of permissible $\delta \cdot \epsilon(\omega)$, $\langle I(r, s) \rangle$ must be monotonically increasing.

This yields:

$$\frac{\mathcal{S}_N(\omega)}{A(\omega) S_0^2(\omega) + B(\omega) S_0(\omega) + \mathcal{S}_N^2(\omega)} = \mu \qquad \text{for all} \quad \omega \quad \text{such that} \quad S_0(\omega) \neq 0, \tag{41}$$

and

$$\frac{1}{\mathcal{S}_N(\omega)} - \mu \leqq 0 \quad \text{for all} \quad \omega \quad \text{such that} \quad S_0(\omega) = 0. \tag{42}$$

(This equation is simply the integrand for $S_0(\omega) = 0$.)

Or from the above for $S_0(\omega) \neq 0$,

$$S_0(\omega) = \frac{-B(\omega) \pm \{B^2(\omega) - 4A(\omega)C(\omega)\}^{\frac{1}{2}}}{2A(\omega)}$$

where

$$C(\omega) = \mathcal{S}_N^2(\omega) - \mathcal{S}_N(\omega)/\mu. \tag{43}$$

Now, $S_0(\omega)$ must be nonnegative. In order that equation (43) yield a nonnegative result, the positive root must be taken, and further the following inequality must be satisfied:

$$\{B^2(\omega) - 4A(\omega)C(\omega)\}^{\frac{1}{2}} \geqq B(\omega). \tag{44}$$

Relation (44) implies

$$C(\omega) \leqq 0 \quad \text{for all } \omega \text{ such that } S_0(\omega) > 0. \tag{45}$$

Relation (42) (which implies $1/\mu \leqq 0$) can be rewritten as

$$C(\omega) \geqq 0 \quad \text{for all } \omega \text{ such that } S_0(\omega) = 0. \tag{46}$$

If we assume that $\mathcal{S}_N(\omega)$ is a monotonically increasing function* of ω, relations (45) and (46) imply

$$\frac{1}{\mu} = \mathcal{S}_N(\omega_{\max}), \tag{47}$$

and

$$S_0(\omega) = 0 \qquad \omega \geqq \omega_{\max}. \tag{48}$$

Hence† rewriting equation (43)

$$S_0(\omega) = \mathcal{S}_N(\omega) \frac{\{1 + 4A(\omega)\,\mathcal{S}_N(\omega_{\max})/\,\mathcal{S}_N(\omega)\}^{\frac{1}{2}} - \{1 + 4A(\omega)\}^{\frac{1}{2}}}{2A(\omega)} \tag{49}$$

where $\omega_{\max}$ can be found from

$$\frac{1}{2\pi}\int_0^{\omega_{\max}} S_0(\omega)\, d\omega = P. \tag{50}$$

This can be used to obtain a relation between $\omega_{\max}$ and P.

---

* This is not at all necessary. It just simplifies the form of the following equations.

† In what follows, it is understood that the expression given for $S_0(\omega)$ holds for $\omega \leqq \omega_{\max}$, and that $S_0(\omega) = 0$ otherwise.

Now if the crosstalk is small [$|X(\omega)| \ll 1$], each of the radicals in the equation for $S_0(\omega)$ can be approximated by the first few terms in the binomial expansion. Then

$$S_0(\omega) = [S_N(\omega_{\max}) - S_N(\omega)]\Big[1 - \tfrac{1}{4}\,|X(\omega)|^2\,(|X(\omega)|^2 + 1)$$

$$+ \tfrac{1}{8}\,|X(\omega)|^2\,(|X(\omega)|^2 + 1)\left(\frac{S_N(\omega)_{\max}}{S_N(\omega)} - 1\right) + \cdots\Big], \quad (51)$$

$$S_0(\omega) \cong [S_N(\omega_{\max}) - S_N(\omega)][1 - \tfrac{1}{4}\,|X(\omega)|^2\,(|X(\omega)|^2 + 1)],$$

$$\cong [S_N(\omega_{\max}) - S_N(\omega)]. \quad (52)$$

This result indicates that for small crosstalk, the signal spectrum ought to be designed independent of the crosstalk spectrum. Equation (52) is the familiar "spectrum filling" result for additive Gaussian noise.* This is shown in Fig. 5. The exact solution [equation (51)] (superimposed

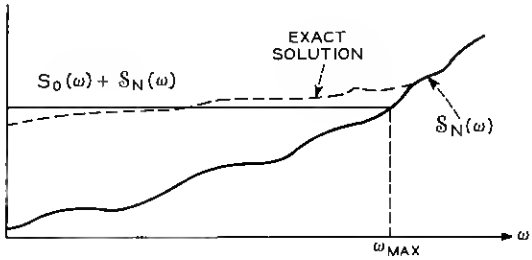

Fig. 5 — Typical optimum spectrum 1.

with the broken lines) simply implies some shaping. Figure 6 illustrates the solution when $S_N(\omega)$ is not monotonically increasing.

With the solution obtained for $S_0(\omega)$

$$\text{Capacity} = \frac{1}{2\pi}\left\{\int_0^{\omega_{\max}} \log\left[1 + \frac{1}{|X(\omega)|^2}\right] d\omega\right.$$

$$\left. + \frac{1}{2}\int_0^{\omega_{\max}} \log\left\{1 + 4\,|X(\omega)|^2\,[|X(\omega)|^2 + 1]\,\frac{S_N(\omega_{\max})}{S_N(\omega)}\right\} d\omega\right\}. \quad (53)$$

Using the small crosstalk approximation for $S_0(\omega)$:

---

* This result is contained in Fano[5], pp. 173ff. That proof is not as rigorous as the one presented here as Fano does not prove that the optimum spectrum is zero whenever the noise is greater than a threshold. Fano's result can be proved directly by noting the discussion preceding equation (21) herein.

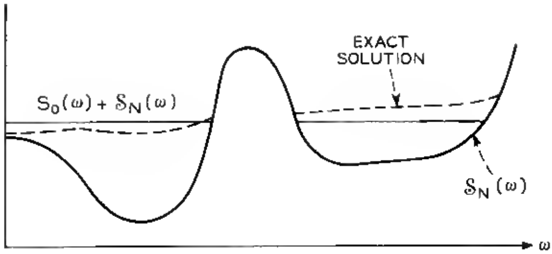

Fig. 6 — Typical optimum spectrum 2.

Capacity

$$\cong \frac{1}{2\pi}\int_0^{\omega_{max}} \log\left[1 + \frac{S_N(\omega_{max}) - S_N(\omega)}{| X^2(\omega) | [S_N(\omega_{max}) - S_N(\omega)] + S_N(\omega)}\right] d\omega,$$

$$\cong \frac{1}{2\pi}\int_0^{\omega_{max}} \log\left[1 + \frac{S_N(\omega_{max}) - S_N(\omega)}{| X(\omega) |^2 S_N(\omega_{max}) + S_N(\omega)}\right] d\omega,$$

$$\cong -\frac{1}{2\pi}\int_0^{\omega_{max}} \log\left[| X(\omega) |^2 + \frac{S_N(\omega)}{S_N(\omega_{max})}\right] d\omega. \qquad (54)$$

(Note that even for small crosstalk, capacity is still a function of the crosstalk spectrum.)

REFERENCES

1. Gallager, R. G., *Information Theory and Reliable Communication,* New York, N. Y.: John Wiley & Sons, Inc., 1968, pp. 369–370.
2. Blackwell, D., Breiman, L., and Thomasian, A. J., "The Capacity of a Class of Channels," *Ann. of Math. Stat., 30,* No. 4, (December 1959), pp. 1229–1241.
3. Pinsker, M. S., "A Quantity of Information of a Gaussian Random Stationary Process, Contained in a Second Process Connected with it in a Stationary Manner," *Doklady Akad. Nauk S.S.S.R., 99,* No. 2, (February 1954), pp. 213–216. (Translation available as A.C.S.I.L. Trans. No. 1266, October 1960).
4. Kolmogorov, A. N., "On the Shannon Theory of Information Transmission in the Case of Continuous Signals," *IRE Trans., IT-2,* No. 4 (December 1956), pp. 102–108.
5. Fano, R. M., *Transmission of Information*, Cambridge, Mass.: The MIT Press, 1961, pp. 173–174.